Artificial Intelligence 175 (2011) 935–937

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

Book review

S. Russell, P. Norvig, Artificial Intelligence: A Modern Approach, Third Edition.

Ashish Sabharwal [a], Bart Selman [b,*]

[a] *IBM Watson Research Center, Yorktown Heights, NY 10598, USA*
[b] *Dept. of Computer Science, Cornell University, Ithaca, NY 14853-7501, USA*

ARTICLE INFO

Article history:
Received 12 January 2011
Accepted 12 January 2011
Available online 19 January 2011

Since the publication of its first edition in 1995, *Artificial Intelligence: A Modern Approach* has become a classic in our field. Even researchers outside AI, working in other areas of computer science, are familiar with the text and have gained a better appreciation of our field thanks to the efforts of its authors, Stuart Russell of UC Berkeley and Peter Norvig of Google Inc. It has been adopted by over 1000 universities in over 100 countries, and has provided an excellent introduction to AI to several hundreds of thousands of students worldwide. The book not only stands out in the way it provides a clear and comprehensive introduction to almost all aspects of AI for a student entering our field, it also provides a tremendous resource for experienced AI researchers interested in a good introduction to subfields of AI outside of their own area of specialization. In fact, many researchers enjoy reading insightful descriptions of their own area, combined, of course, with the tense moment of checking the author index to see whether their own work made it into the book. Fortunately, due in part to the comprehensive nature of the text, almost all AI researchers who have been around for a few years can be proud to see their own work cited. Writing such a high-quality and erudite overview of our field, while distilling key aspects of literally thousands of research papers, is a daunting task that requires a unique talent; the Russell and Norvig author team has clearly handled this challenge exceptionally well.

Given the impact of the first edition of the book, a challenge for the authors was to keep such a unique text up-to-date in the face of rapid developments in AI over the past decade and a half. Fortunately, the authors have succeeded admirably in this challenge by bringing out a second edition in 2003 and now a third edition in 2010. Each of these new editions involves major rewrites and additions to the book to keep it fully current. The revisions also provide an insightful overview of the evolution of AI in recent years.

The text covers essentially all major areas of AI, while providing ample and balanced coverage of each of the sub-areas. For certain subfields, part of the text was provided by respective subject experts. In particular, Jitendra Malik and David Forsyth contributed the chapter on computer vision, Sebastian Thrun wrote the chapter on robotics, and Vibhu Mittal helped with the chapter on natural language. Nick Hay, Mehran Sahami, and Ernest Davis contributed to the engaging set of exercises for students. Overall, this book brings together deep knowledge of various facets of AI, from the authors as well as from many experts in various subfields. The topics covered in the book are woven together via the theme of a grand challenge in AI — that of creating an *intelligent agent*, one that takes "the best possible [rational] action in [any] situation." Every aspect of AI is considered in the context of such an agent. For instance, the book discusses agents that solve problems through search, planning, and reasoning, agents that are situated in the physical world, agents that learn from observations, agents that interact with the world through vision and perception, and agents that manipulate the physical world through

* Corresponding author.
E-mail addresses: ashish.sabharwal@us.ibm.com (A. Sabharwal), selman@cs.cornell.edu (B. Selman).

doi:10.1016/j.artint.2011.01.005

electronic or robotic controls. This unifying theme goes a long way in tying together the vast variety of specialized subfields of AI.

The third edition of the book is divided into seven main parts: foundations of AI and intelligent agents; problem-solving through search; knowledge representation, reasoning, and planning; dealing with uncertainty; learning; perception and action in the physical world through natural language, vision, and robotics; and philosophical issues surrounding AI. Each part consists of typically four to six chapters, each with roughly half a dozen sections. All in all, the 1132 pages of this text include 27 chapters, 177 sections, 416 figures, nearly 2000 citations, and a 38 page long index consisting of almost 6000 items. Literally an intellectual tour de force. Although the various parts of the book can be studied in isolation, a key factor behind the success of this book is that it helps the reader realize the multitude of intricate connections between various subareas of AI. When used in the classroom, students taking a one semester "Introduction to AI" course would benefit the most from touching upon each of the seven main parts of the book but limiting the overall amount of material by focusing on only a few chapters within each part.

In addition to clearly articulated technical material, the book also provides towards the end of every chapter a historical perspective, researched in part by Douglas Edwards, and a brief literature overview. These perspectives and citations provide a highly useful and engaging context for both students and AI researchers. The plentiful examples and figures are invaluable to the reader. Most topics are introduced from the ground up, covering derivations of nearly all foundational concepts and results in AI. To mention just a few examples, the text explains the optimality of A*, the completeness of resolution, the graphplan framework for planning, inference in Bayesian networks, particle filtering algorithms, back-propagation, PAC learning, the EM algorithm, support vector machines, etc. It also discusses historical ideas that lie at the heart of computation and AI, such as the Turing test, completeness and incompleteness, and undecidability. One can also find in this book amazingly clear and concise introductions to a wide range of core topics and concepts in computer science and even other related areas of science. Such topics include computational complexity, logic, formal languages and grammars, probability theory, control theory, game theory, decision theory, and notions from psychology and philosophy. The sheer breadth of the topics the text covers illustrates the tremendous richness and diversity of our field, which is a large part of its appeal as a field of study.

For the second edition, which appeared in 2003, Russell and Norvig expanded the text in a variety of areas including: historical connections of AI to fields such as control theory, game theory, economics, and neuroscience; additional coverage of online algorithms, constraint satisfaction problems, and modern satisfiability (SAT) solvers; new planning approaches including graph-based and SAT-based planning; probabilistic inference techniques such as for Bayesian inference, dynamic Bayesian networks, hidden Markov models, and Kalman filters; and ethical implications of AI in human society. This constituted a substantial expansion of the first edition, providing good coverage of many significant contributions to the theory and practice of AI during the late 1990s and early 2000s.

For the third edition of the text, Russell and Norvig again went back to the drawing board to incorporate the key developments in our field in the past seven to ten years. Nearly 20% of the citations in this edition are to works published after the printing of the second edition. Probabilistic reasoning, statistical methods, and multi-agent approaches, in particular, have grown yet further in importance and the new edition appropriately reflects this development. There is new emphasis on the challenge of unifying probabilistic and logical reasoning and representations. Expanding upon the underlying theme of representation by an intelligent agent, the authors now distinguish between atomic, factored, and structured representations, highlighting trade-offs in terms of relevance, usefulness, and computational efficiency of working with these representations in various contexts. This edition also provides a more comprehensive introduction to the area of machine learning, which has witnessed tremendous growth over the past two decades, both in terms of theoretical developments and practical impact. Partial observability has also received more attention in this edition in various contexts, such as in search, planning, and game playing. Needless to say, sections such as 'state-of-the-art game programs' have been updated to reflect the latest in the respective subfields. Finally, in keeping with the ever growing role vast amounts of information plays in our lives, this edition includes expanded coverage of Web search engines, techniques for information extraction, and methods to work with extremely large data sets. In terms of exposition, hundreds of new exercises have been added, along with substantial new material on the book's website, http://aima.cs.berkeley.edu, including sample source code, lecture slides, and suggested syllabi.

As the field of AI continues to grow and expand at a rapid pace, there is a significant pressure to specialize within the field. Most universities now offer a range of courses in various subfields of AI, such as machine learning, natural language processing, graphical models, knowledge representation, computer vision, robotics, etc. There are also a number of good textbooks written specifically for these subfields. Relying solely on these specialized courses and books, however, poses the risk that new students entering the field of AI may no longer get a good overview of the richness and complexity of the diverse research challenges in AI. Moreover, many themes of AI, such as trade-offs between expressiveness of representations and their computational properties, for instance between propositional and first-order representations, cut across different subfields of AI. We strongly believe that it is important for students to be aware of these general themes underlying AI problems and to get a solid introduction to the broad range of topics in AI before embarking on a more specialized subfield. In fact, it is likely that some of the future breakthroughs in AI will come from students taking concepts from different parts of Russell and Norvig's text and combining them in novel ways.

In summary, this true AI classic remains an exceptional resource for both new students and seasoned AI researchers wishing to learn about a subfield of AI or to get a good understanding of the overarching scope of AI as an integral part of

computer science. Over the years, the text will continue to educate in a superb manner literally hundreds of thousands of computer scientists about the achievements and the excitement of our field. The authors' ability to distill the essence of AI concepts and techniques as well as their rich inter-connections is simply unsurpassed. Moreover, the revisions — going from the first edition in 1995 to the second edition in 2003 and now the third edition in 2010 — beautifully keep pace with our fast moving and continuously expanding field.